Trading Strategies using R
The quest for the holy grail

Eran Raviv
Econometric Institute - Erasmus University,
http://eranraviv.com

April 02, 2012
Outline for section 1

1. introduction
2. Connection and data
3. The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading
4. Final Comments
The strategy does not even cover transaction costs. Buy and hold is much better for the period measured.

Report for the period 07/2009 - 07/2010

-0.15 -0.05 0 0.05 0.1 0.15 0.2
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97 101 105 109
Cumulative Returns
cumsum
Decision was made to reduce volume

Mean Return 0.000046
Cumulative Return 0.006818323
Proportion of Success 0.472
Total volume 6,039,540
Average Time held 3 hours
# of Trades 531
Average number of trades (day) 3.6
# Trading Days 147
Commission paid for trades 1638
Commission paid for Data 462
SnP return for period 12.00%
Outline for section 2

1. introduction
2. Connection and data
3. The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading
4. Final Comments
For Inter-day, \textit{yahoo} is fine

\begin{verbatim}

nam = c('AON', 'MMC', 'AKS', 'BAC', ...) ; tckr = sort(nam)
# Most recent 252 days:
end<- format(Sys.Date(),"%Y-%m-%d") # yyyy-mm-dd
start<-format(Sys.Date() - 365,"%Y-%m-%d")
l = length(tckr)
dat = array(dim = c(252,6,l))
for (i in 1:l){
dat0 = (getSymbols(tckr[i], src="yahoo", from=start, to=end,
auto.assign = FALSE)) # Cancel auto.assign if you want to
    manipulate the object
dat[1:length(dat0[,2]),,i] = dat0[,2:6]
}
dat = dat[1:length(na.omit(dat[,1,1])),,]

\end{verbatim}

Eran Raviv
Trading Strategies using R
April 02, 2012
For Intra-day use *IB*

- IB has extensive API. Connect to their trading platform (TWS) using *Java* and *C* among others.
For Intra-day use *IB*

- IB has extensive API. Connect to their trading platform (TWS) using *Java* and *C* among others.
IB has extensive API. Connect to their trading platform (TWS) using *Java* and *C* among others.

Account is not that easy to set up, many forms to fill out and hefty sum to transfer, especially if you would like to day trade.
For Intra-day use *IB*

- IB has extensive API. Connect to their trading platform (TWS) using *Java* and *C* among others.

- Account is not that easy to set up, many forms to fill out and hefty sum to transfer, especially if you would like to day trade.
IB has extensive API. Connect to their trading platform (TWS) using *Java* and *C* among others.

Account is not that easy to set up, many forms to fill out and hefty sum to transfer, especially if you would like to day trade.

**Jeffrey A. Ryan** did outstanding work, we can now trade via *R*. 
For Intra-day use IB

- Easy:

```r
library(IBrokers)
IBrokers version 0.9-1: Implementing API Version 9.64
This software comes with NO WARRANTY. Not intended for production use! See ?IBrokers for details
con = twsConnect(clientId = 1, host = 'localhost', port = 7496, verbose = TRUE, timeout = 5, filename = NULL)
```
For Intra-day use **IB**

- Easy:
  
  ```r
  library(IBrokers)
  IBrokers version 0.9-1: Implementing API Version 9.64
  This software comes with NO WARRANTY. Not intended for production use! See ?IBrokers for details}
  con = twsConnect(clientId = 1, host = 'localhost', port = 7496, verbose = TRUE, timeout = 5, filename = NULL)
  ```

- High frequency data if you have the patience to program it.
For Intra-day use **IB**

- **Easy:**

```
library(IBrokers)
IBrokers version 0.9-1: Implementing API Version 9.64
This software comes with NO WARRANTY. Not intended for production use! See ?IBrokers for details
con = twsConnect(clientId = 1, host = 'localhost', port = 7496, verbose = TRUE, timeout = 5, filename = NULL)
```

- High frequency data if you have the patience to program it.
- Limitation on the number of requests.
For Intra-day use IB

- Easy:
  ```r
  library(IBrokers)
  IBrokers version 0.9-1: Implementing API Version 9.64
  This software comes with NO WARRANTY. Not intended for production use! See ?IBrokers for details
  ```
  ```r
  con = twsConnect(clientId = 1, host = 'localhost', port = 7496, verbose = TRUE, timeout = 5, filename = NULL)
  ```

- High frequency data if you have the patience to program it.
- Limitation on the number of requests.
- In any case not more than one year, but you can store it.
For Intra-day use **IB**

- **Easy:**
  ```r
  library(IBrokers)
  IBrokers version 0.9-1: Implementing API Version 9.64
  This software comes with NO WARRANTY. Not intended for production use! See ?IBrokers for details}
  con = twsConnect(clientId = 1, host = 'localhost', port = 7496, verbose = TRUE, timeout = 5, filename = NULL)
  ```

- High frequency data if you have the patience to program it.
- Limitation on the number of requests.
- In any case not more than one year, but you can store it.
- Professional yahoo group at:
  ```
  http://finance.groups.yahoo.com/group/TWSAPI/
  ```
Outline for section 3

1. Introduction
2. Connection and data
3. The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading
4. Final Comments
Selected Ideas

Over the years I have backtested many ideas, among others:

- Sign Prediction
Selected Ideas

Over the years I have backtested many ideas, among others:

- Sign Prediction
- Filtering
Selected Ideas

Over the years I have backtested many ideas, among others:

- Sign Prediction
- Filtering
- Multivariate time series modelling
Over the years I have backtested many ideas, among others:

- Sign Prediction
- Filtering
- Multivariate time series modelling
- Pairs trading
Over the years I have backtested many ideas, among others:

- Sign Prediction
- Filtering
- Multivariate time series modelling
- Pairs trading
Over the years I have backtested many ideas, among others:

- Sign Prediction
- Filtering
- Multivariate time series modelling
- Pairs trading

Born to trade, forced to work.
Table of Contents

1 introduction

2 Connection and data

3 The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading

4 Final Comments
Sign prediction using:

- Logistic Regression (*glm*)
- Support Vector Machine (*svm*)
  - `library(e1071)`
- K-Nearest Neighbour (*knn*)
  - `library(class)`
- Neural Networks (*nnet*)
  - `library(nnet)`
Working with **daily returns**, so target is to predict tomorrow’s move. (Avoid overnight)

Explanatory variables considered:

- **I** five lags (one week)
- **II** Spread between the volume and the rolling average of most recent 5 days.
- **III** Volatility - average of the last five days.
Volatility is measured as the average of three different intra-day volatility measures which are more efficient (converge faster) than the standard ”sd” estimate:

- **Parkinson (1980):**
  \[
  \sigma = \sqrt{\frac{1}{4N\ln 2} \sum_{i=1}^{N} (\ln \frac{h_i}{l_i})^2}
  \]
Volatility is measured as the average of three different intra-day volatility measures which are more efficient (converge faster) than the standard "sd" estimate:

- Parkinson (1980):
  \[ \sigma = \sqrt{\frac{1}{4N\ln 2} \sum_{i=1}^{N} (\ln h_i^l_i)^2} \]

- German Klass (1980):
  \[ \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} (\ln h_i^l_i)^2 - \frac{1}{N} \sum_{i=1}^{N} (2\ln 2 - 1)(\ln \frac{c_i}{c_{i-1}})^2} \]
Volatility is measured as the average of three different intra-day volatility measures which are more efficient (converge faster) than the standard ”sd” estimate:

- Parkinson (1980):
  \[ \sigma = \sqrt{\frac{1}{4Nln2} \sum_{i=1}^{N} (ln \frac{h_i}{l_i})^2} \]

- German Klass (1980):
  \[ \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} (ln \frac{h_i}{l_i})^2 - \frac{1}{N} \sum_{i=1}^{N} (2ln2 - 1)(ln \frac{c_i}{c_{i-1}})^2} \]

- Rogers and Satchell (1991):
  \[ \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (ln \frac{h_i}{l_i})(ln \frac{h_i}{o_i}) + (ln \frac{l_i}{c_i})(ln \frac{l_i}{o_i})} \]
Sign Prediction - continued

dat0 = (getSymbols(tckr[1], src="yahoo", from=start, to=end, auto.assign = FALSE))
l = length(dat0[,1])
dates0 = (index(dat0)) # trick to get trading dates
tt = NULL ## we now parse it into IB mode
for (i in 1:l){
tt[i] = paste(substr(dates0[i],1,4), substr(dates0[i],6,7),
            substr(dates0[i],9,10), sep = "")
tt[i] = paste(tt[i], "23:00:00 GMT")
}
cont=twsEquity('plug your favourite symbol', 'SMART', 'NYSE')
mat1 = array(dim = c(l,400,8))#Typical day should have 390 mins
for (i in 1:l){
    m1 = as.matrix(reqHistoricalData(con, cont, tt[i], barSize = "1 min",
                     duration = "1 d", useRTH = "1", whatToShow = "TRADES",
                     format = "1", verbose = TRUE))
    mat1[i,1:length(m1[,1]),] = m1
    Sys.sleep(14) ## IB restriction, WAIT.
}
Sign Prediction - continued

Sample code:

```r
logit1 = glm(y~lagy+volat+volume, data=dat[1:t1,], family=
  binomial(link = "logit"), na.action=na.pass)
summary(logit1) #t1 is end of training, TT is full length.
library(nnet)
nnet1 = nnet(as.factor(y)~lagy+volat+volume, data=dat[1:t1,],
  size=1, trace=T)
summary(nnet1)
library(class)
knn1 = knn(dat[1:t1,], dat[(t1+1):TT,], cl = dat$y[1:t1], k=25,
  prob=F)
sum(knn1==dat$y[(t1+1)])/ (TT−t1+1)#Hit ratio
library(e1071)
svm1 = svm(dat[1:t1,2:4], y=dat[1:t1,1], type = "C")
# In sample:
sum(svm1$fit==dat$y[(1):t1])/t1
# out of sample:
svmpred=predict(svm1, newdata = dat[(t1+1):TT,2:4])
sum(svmpred==dat$y[(t1+1):TT])/(TT−t1+1)#Hit ratio
```
Table of Contents

1. introduction
2. Connection and data
3. The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading
4. Final Comments
Deviation from the mean

- Motivation $\rightarrow$ Disposition effect, the \textit{Voodoo} of financial markets.
- Standardise the deviation from the (rolling) mean.
Deviation from the mean

Google

Histogram for Z

Days

Price

Frequency

Z

Days

Eran Raviv
1. introduction

2. Connection and data

3. The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading

4. Final Comments
Motivation

- Momentum in Microstructure - Dermot Murphy and Ramabhadran S. Thirumalai (Job Market Paper - 2011)

Motivation

- Momentum in Microstructure - Dermot Murphy and Ramabhadran S. Thirumalai (Job Market Paper - 2011)


- We find predictable patterns in stock returns. Stocks whose relative returns are high in a given half-hour interval today exhibit similar outperformance in the same half-hour period on subsequent days. The effect is stronger at the beginning and end of the trading day. These results suggest...
For each day \( t = \{1, \ldots, T\} \), the return of half an hour \( k = \{1, \ldots, 13\} \), and the lag number \( p = \{1, \ldots, P\} \):

\[
\begin{bmatrix}
  y_{1,t} \\
  y_{2,t} \\
  \vdots \\
  y_{k,t}
\end{bmatrix}
= \begin{bmatrix}
  c_1 \\
  c_2 \\
  \vdots \\
  c_k
\end{bmatrix}
+ \begin{bmatrix}
  a_{1,1}^1 & a_{1,2}^1 & \cdots & a_{1,k}^1 \\
  a_{2,1}^1 & a_{2,2}^1 & \cdots & a_{2,k}^1 \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{k,1}^1 & a_{k,2}^1 & \cdots & a_{k,k}^1
\end{bmatrix}
\begin{bmatrix}
  y_{1,t-1} \\
  y_{2,t-1} \\
  \vdots \\
  y_{k,t-1}
\end{bmatrix}
+ \cdots +
\begin{bmatrix}
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots
\end{bmatrix}
\begin{bmatrix}
  e_{1,t} \\
  e_{2,t} \\
  \vdots \\
  e_{k,t}
\end{bmatrix}
\]
For each day $t = \{1, \ldots, T\}$, the return of half an hour $k = \{1, \ldots, 13\}$, and the lag number $p = \{1, \ldots, P\}$:

$$
\begin{bmatrix}
y_{1,t} \\
y_{2,t} \\
\vdots \\
y_{k,t}
\end{bmatrix}
= 
\begin{bmatrix}
c_1 \\
c_2 \\
\vdots \\
c_k
\end{bmatrix}
+ 
\begin{bmatrix}
a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\
a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\
\vdots & \vdots & \ddots & \vdots \\
a_{k,1} & a_{k,2} & \cdots & a_{k,k}
\end{bmatrix}
\begin{bmatrix}
y_{1,t-1} \\
y_{2,t-1} \\
\vdots \\
y_{k,t-1}
\end{bmatrix}
+ \cdots +
\begin{bmatrix}
a_{1,1}^p & a_{1,2}^p & \cdots & a_{1,k}^p \\
a_{2,1}^p & a_{2,2}^p & \cdots & a_{2,k}^p \\
\vdots & \vdots & \ddots & \vdots \\
a_{k,1}^p & a_{k,2}^p & \cdots & a_{k,k}^p
\end{bmatrix}
\begin{bmatrix}
y_{1,t-p} \\
y_{2,t-p} \\
\vdots \\
y_{k,t-p}
\end{bmatrix}
+ \begin{bmatrix}
e_{1,t} \\
e_{2,t} \\
\vdots \\
e_{k,t}
\end{bmatrix}
$$

**Problem:** for $P = 1$, how many parameters?
VAR models (cont’d)

- Possible solution $\implies$ Dimension Reduction.
VAR models (cont’d)

• Possible solution $\implies$ Dimension Reduction.

• Stepwise Regression, Lasso, Variable selection (according to some Information Criteria), Principal Component Regression, Ridge Regression, Bayesian VAR and many more.

• Very nice vars package to start you off, though as most built-ins, not flexible enough. (e.g. rolling windows and/or shrinking)
We will now talk about pairs trading.

- Well known and widely used. (e.g. *Statistical Arbitrage in the U.S. Equities Market*, Marco Avellaneda and Jeong-Hyun Lee (2008))
Pairs Trading

- Well known and widely used. (e.g. *Statistical Arbitrage in the U.S. Equities Market*, Marco Avellaneda and Jeong-Hyun Lee (2008))

- Suitable for the conservative mind. (we see why in a minute..)
Well known and widely used. (e.g. *Statistical Arbitrage in the U.S. Equities Market*, Marco Avellaneda and Jeong-Hyun Lee (2008))

Suitable for the conservative mind. (we see why in a minute..)
The Idea:

\[
\begin{align*}
    r_a &= \beta_a r_m + e_a \\
    r_b &= \beta_b r_m + e_b \\
    r_{ab} &= w_a (\beta_a r_m + e_a) + w_b (\beta_b r_m + e_a) \\
    &= r_m (w_a \beta_a + w_b \beta_b) + \text{noise}
\end{align*}
\]

and so with weights \( w_a = -\frac{\beta_b}{\beta_a - \beta_b} \) and \( w_b = 1 - w_a \) we can net out the market. (and other factors if you will)
Choose symbols with similar properties.

Net out the market and create the spread:

```r
## sp1 = stock price 1, g=size of moving window,
## n = length(sp1)
for (i in g:n){
  bet0[i]=lm(sp1[(i-g+1):(i-1)] ~ sp2[(i-g+1):(i-1)])**coeffet[1]  #
  # note -> i-1
  bet1[i]=lm(sp1[(i-g+1):(i-1)] ~ sp2[(i-g+1):(i-1)])**coeffet[2]
  spread[,]=sp1[(i-g+1):i]-rep(bet0[i],g)-bet1[i]*sp2[(i-g+1):i]
}
```

Text book example (actually from: *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*)
Choose symbols with similar properties.

Net out the market and create the spread:

```r
## sp1 = stock price 1, g=size of moving window,
## n = length(sp1)
for (i in g:n){
  bet0[i] = lm(sp1[(i-g+1):(i-1)] ~ sp2[(i-g+1):(i-1)])$coef[1]  #
  # note -> i-1
  bet1[i] = lm(sp1[(i-g+1):(i-1)] ~ sp2[(i-g+1):(i-1)])$coef[2]
  spread[, i] = sp1[(i-g+1):i] - rep(bet0[i], g) - bet1[i] * sp2[(i-g+1):i]
}
```

Text book example (actually from: *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*)

The GLD and GDX spread
The GLD and GDX spread:

Last 30 days

Last 90 days

Last 180 days

Last 365 days

Eran Raviv
Trading Strategies using R
April 02, 2012
Estimation of the market neutral portfolio is tricky:

- Price levels or price changes?
Estimation of the market neutral portfolio is tricky:

- Price levels or price changes?

- Stability over time
Estimation of the market neutral portfolio is tricky:

- Price levels or price changes?
- Stability over time
- Errors on both sides. (both $y$ and $x$ are measured with errors)
Pairs trading issues

Stability over time:

- Beta for the last 30 days = 2.1
- Beta for the last 90 days = 1.77
- Beta for the last 180 days = 2.26
- Beta for the last 365 days = 1.82
Errors on both sides:

\[ st_a = \alpha st_b + e_a \]
\[ st_b = \beta st_a + e_b \]

\[ \hat{\alpha} \neq \frac{1}{\hat{\beta}} \]

Portfolio is different and will depend on which instrument goes on the LHS and which on the RHS.
Pairs trading - possible solutions

Price levels or price changes?
Pairs trading - possible solutions

Price levels or price changes?

- flip a coin (solid option)
- average the estimates
Pairs trading - possible solutions

Price levels or price changes?

- flip a coin (solid option)
- average the estimates
Stability over time
Stability over time

- Choose window length that fits your style, the shorter the more you trade.
- Recent paper (though in different context) suggests to average estimates across different windows to partially hedge out uncertainty. *(M. Hashem Pesaran, Andreas Pick. Journal of Business and Economic Statistics. April 1, 2011)*
- Kalman filter the coefficients.
Errors on both sides, two highly correlated possible solutions:

- Demming regression (1943). (Total least squares - just minimize numerically both sides simultaneously)
- Geometric Mean Regression - force coherence through:

\[
\begin{align*}
st_a &= \alpha st_b + e_a \\
st_b &= \beta st_a + e_b \\
\hat{\gamma} &= \sqrt{\hat{\alpha} \times \frac{1}{\hat{\beta}}}
\end{align*}
\]
Outline for section 4

1. introduction
2. Connection and data
3. The quest
   - Sign Prediction
   - Filtering
   - Time Series Analysis
   - Pairs Trading
4. Final Comments
Miscellaneous remarks

- Trading costs!, consider it when backtesting.
Miscellaneous remarks

- Trading costs!, consider it when backtesting.
- You cannot be too careful, stay pessimistic.
Miscellaneous remarks

- Trading costs!, consider it when backtesting.
- You cannot be too careful, stay pessimistic.
- Adopt rigorous robustness checks, different instruments, different time frames and even different markets.
Miscellaneous remarks

- Trading costs!, consider it when backtesting.
- You cannot be too careful, stay pessimistic.
- Adopt rigorous robustness checks, different instruments, different time frames and even different markets.
- Use paper money for at least a full quarter, it will help you handle operational problems. (e.g. outages and time zones issues)
Miscellaneous remarks

- Trading costs!, consider it when backtesting.
- You cannot be too careful, stay pessimistic.
- Adopt rigorous robustness checks, different instruments, different time frames and even different markets.
- Use paper money for at least a full quarter, it will help you handle operational problems. (e.g. outages and time zones issues)
- It is (very) stressing work, know it before you start.
Miscellaneous remarks

- Trading costs!, consider it when backtesting.
- You cannot be too careful, stay pessimistic.
- Adopt rigorous robustness checks, different instruments, different time frames and even different markets.
- Use paper money for at least a full quarter, it will help you handle operational problems. (e.g. outages and time zones issues)
- It is (very) stressing work, know it before you start.
- Know what you are doing, what is your edge? why it is (not) there?
THANKS

and good luck at the tables..